DOCUMENT RESUME

ED 044 438 TM 000 164

AUTHOR	Holtzman, Wayne H.
TITLE	The Changing World of Mental Measurement and Its Social Significance.
PUB DATE	Sep 70
NOTE	20p.; Presidential Address (Division 5) at the annual meeting of the American Psychological Association, Miami Beach, Florida, September 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.10
DESCRIPTORS	Aptitude Tests, Computers, Educational Change, *Educational Equality, *Educational Innovation, *Evaluation, Measurement, *Measurement Techniques, National Norms, Objective Tests, Social Change, Standardized Tests, *Technological Advancement
IDENTIFIERS	National Assessment of Educational Progress

ABSTRACT

Prior to the late fifties test usage enjoyed a degree of acceptance which diminished as test impartiality was increasingly questioned. Criticisms of testing are delineated, including the discrimination implicit in normative testing itself and the application of resulting test scores. This criticism of measurement techniques has directed attention to other inequities in the educational system. Thus, the emergence of new educational techniques and related measurement techniques is a major force in educational reform, resulting in such innovations as new instructional techniques and curricula. Another important departure from standarized normative measures grows out of the increased concern for developing a national system of social indicators. One of the most significant changes in the field of mental measurement in recent years is a recognition of social, cultural, and linguistic variability. In conclusion, it is noted that the electronic computer is necessary to the implementation of most of the new developments in measurement. A bibliography is included. (Author/PR)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED044438

THE CHANGING WORLD OF MENTAL MEASUREMENT AND ITS SOCIAL SIGNIFICANCE[1]

Wayne H. Holtzman

The University of Texas

One of the great success stories of modern psychology is the development of objective tests for measuring human abilities that are of importance to society. During the past half century the standardized mental test with nationally based norms has proven to be a highly effective instrument for selection and classification of men in the armed forces, for evaluation of educational progress within our school systems, for selective admission of college students, for selection of employees within government, business, and industry, and for clinical assessment of individuals in need of psychological services. It is estimated that within American schools alone, over 250 million standardized tests of ability are administered each year. (Brim, et al, 1969). It is a rare individual indeed, especially among children and young adults, who has not been evaluated by a standardized mental test, a test that has played a significant role in determining his place in society.

From World War I until the late fifties, the testing movement enjoyed a degree of public acceptance it is unlikely to see again. Judging each person on the basis of his measured performance rather than on his family background, social status, or political connections has been a powerful

TM 000 164

---

[1] Presidential Address of Division 5, presented at the annual meeting of the American Psychological Association, Miami Beach, Florida, September, 1970.

agent of social change. Assuming unbiassed, reliable measurement, what could be more just within the American concept of an egalitarian society than recognizing merit by objective tests of ability? Even today, college entrance examinations have made it possible for able but financially poor students to obtain scholarships in the best private colleges.

Criticisms of Testing

By the late fifties it became generally apparent that the large-scale normative use of objective tests for rewarding selected individuals among many in competition has serious social consequences of debatable value. The testing movement has always had its critics but they failed to gain a foothold until the impact of adverse decisions based on tests had been felt by millions of individuals. In the post-Sputnick period, a growing number of critics have claimed that mental tests are unfair to the bright but unorthodox person, to the culturally disadvantaged, and to the naive individual who lacks experience in taking standardized tests (Anastasi, 1967; College Entrance Examination Board, 1970).

The growing controversies surrounding mental tests have become especially acute within educational institutions. It is generally recognized that the educated person enjoys the riches of society as well as enhanced self-esteem and personal development, while the person who prematurely drops out of school is cast into an inferior role. It is not surprising that the angry cries of black students are directed at normative tests which deprive them of entrance to the better colleges, jobs, and social positions.

A major dilemma arises in attempting to meet these criticisms. The traditional academic curricula of our schools and colleges are becoming

increasingly dependent upon verbal communication, verbal memory, and the same kind of abstract reasoning as measured by scholastic aptitude tests. Therefore, sufficiently high correlations arise between standardized multiple-choice aptitude tests and course examinations to justify the use of tests for prediction of academic achievement and selective admissions. The rapid growth of higher education and the greatly increased number of students per course has forced more and more instructors to employ multiple-choice objective examinations for grading students. As a result, the relevance of scholastic aptitude tests for prediction of academic grades has increased, rather than decreased, in recent years. The compelling economics of mass education and objective normative testing are exceedingly difficult to resist in a rapidly expanding system of higher education. Tests that are designed for normative use, whether for college admissions or course examinations, discriminate against those who are culturally different from the majority.

Such incidental discrimination might be more justifiable if there were a close correspondence between success in school and subsequent occupational success. But for a number of reasons, the correlation between grades and later success is too low to argue generally that measured performance in the traditional academic curriculum is that critical. The issue is made more complex by the fact that entry to many occupations is denied an individual who fails to complete the prescribed academic program, regardless of the program's relevance. The growing meritocracy built around traditional curricula that are uniformly prescribed, normative tests that are competitively graded, and restrictive credentials for job entry may be an efficient means of building a

technological society, but it does so by exacting a heavy toll upon those members of society who fail to conform to the majority. The more tightly the meritocracy is drawn, the more self-fulfilling the prophecies.

Educational Reform and the New Technologies

A way out of this dilemma may be closer at hand than many of us realize. The number of pressures within American society and new developments in measurement and instruction are moving in the same general direction. Led by students, spokesmen for minority rights, and concerned academicians, the general public is becoming increasingly aware of serious inequities within our educational system. As higher education becomes more essential to vocational advancement and personal fulfillment, the fruits of education cannot be denied to anyone who is motivated and capable of profiting from it.

The growing attacks upon normative testing for college admission and course grading are having an impact as more and more individuals call for less emphasis upon scholastic aptitude measures and more upon other abilities and new forms of instruction. The kinds and variety of curricula recognized as appropriate for various forms of education are increasing markedly. Courses aimed at social problems and individual self-development are eroding the traditional, discipline-oriented curricula in many colleges. This new thrust may involve individual competencies in such things as social leadership, self-awareness, regard for human rights and social responsibilities or other aspects of behavior which typically have not been important in traditional academic pursuits. As the curriculum moves through reform there will be opportunities for new kinds of measurement as well.

Emphasis is being given in many circles to the idea of individualized instruction in which the learner moves at his own pace and at a time and place that is appropriate for him as an individual. The units of instruction emphasize self-paced learning with regular social reinforcement to maintain a high degree of motivation and relevance, coupled with the concept of continuous progress from one unit to the next. These "micro-curriculum units" or modules have fairly well defined behavioral objectives or performance criteria by which mastery can be recognized. The curriculum itself is viewed in a more global manner as consisting of strings of modules arranged accoring to an explicit hierarchy of values that are in harmony with the future goals of individual development. In many fields of learning these specific modules involve training objectives where criterion testing for standardized mastery is employed rather than normative testing for measuring individual differences. Much of what goes on in education is susceptible to treatment in this form. The broader educational objectives differ considerably from one individual to the next in order to maximize potentiality for individual development.

A major force for social change in educational reform is the emergence of new educational technology and related techniques of measurement. Keeping track of a person moving at his own pace in a continuous progress environment where the particular branching of the curriculum is tailor-made for the student's own learning aptitudes and level, requires a computer to manage the curriculum and assist with the instruction (Holtzman, 1970). In a traditional setting, the instructor keeps a record of how well each student does on each achievement test for the course, while the periodically collected scores from standardized normative tests are stored

centrally. When instruction is individualized, testing must be done more frequently and at different times for each student. In many cases performance testing and instruction are so closely interwoven that they appear as one integrated learning activity. Except for periodic testing at a later date to determine how much a person has retained, even the conceptual nature of measurement shifts from a normative basis, where each person is compared with a general population, to a criterion-referenced basis, where the only decision made is whether or not the student has achieved the desired objective for a specific instructional module. Not only are more short tests given but many more have to be constructed, again requiring a computer for generating tests from item pools as well as scoring and storing them for each student.

Several large-scale programs of individualized instruction are sufficiently advanced to demonstrate the feasibility and power of this approach to educational reform. Now in its fourth year of operation under the leadership of John Flanagan and jointly developed by the American Institutes for Research and Westinghouse Learning Corporation, Project PLAN consists of over a thousand modules divided across nine operating grades and four subject-matter areas (Dunn, 1969). Each teaching unit is coded as to reading difficulty, required teacher supervision, media richness, required social involvement, and a number of other characteristics. A profile is prepared for each student containing measures of abilities, interests, aspirations, and background data for use by the computer in matching the curriculum to the student. The combination of normative measurement on nationally standardized tests for initial guidance and placement of the student and criterion-referenced tests for assessing progress in mastering the curriculum modules is especially noteworthy.

Experience to date with over ten thousand students indicates that most individuals like the new freedom provided by PLAN, and that learning proceeds at a faster pace.

A still more detailed form of individualized instruction can be found in the program of Individually Prescribed Instruction developed by Glaser and associates at the University of Pittsburgh's Learning Research and Development Center (Cooley and Glaser, 1969). A specific lesson plan is prescribed individually for each child every day, depending upon his performance and desires of the previous day. Thousands of curriculum modules are stored and retrieved manually by clerks at the end of each day until the experimental system can be perfected and stored electronically in computers. Interwoven with each module is a criterion-referenced achievement test that provides a basis for decision-making in selecting the next module.

A recent study by Ferguson (1968) serves to illustrate computer-assisted branched testing with elementary arithmetic materials in the Pittsburgh IPI program. A model was developed and tested in which items are selected on the basis of previous responses and are thus tailored to the competencies of the student. A learning hierarchy of prerequisite relationships among eighteen objectives in addition and subtraction was formulated on the basis of previous studies. Two major sequences emerged as dominant in the hierarchy, one involving only addition skills and the other exclusively concerned with subtraction. A third sequence integrated both addition and subtraction. Initially, an examinee was presented with a randomly-generated item for the specific objective being tested. The computer scored his response as correct or incorrect and

generated another item. The process continued until a sufficient number of items had been answered for the computer to make a decision regarding the individual's proficiency on the objective. The decision model involved assigning a priori probability values to the two types of error constituting incorrect decisions and applying Wald's sequential probability ratio test to terminate the testing on the objective in question. Selection of the next objective to be tested depended upon the examinee's proficiency on the first objective as well as the proposed learning hierarchy. When given to 75 students in grades one through six at the Oakleaf Elementary School, the sequential branched testing method proved to be three times as efficient as a fixed-length conventional test, requiring on the average only 52 items instead of 150.

A sequential branched-testing procedure proves far superior to conventional testing when one has a computer for generating and scoring items, a suitable communication terminal for interaction of computer and examinee, and a good basis for arranging the skills to be tested in a learning hierarchy. The procedure is ideally suited to criterion-referenced testing but is of questionable value where normative testing is employed. As Lord (1970) has demonstrated, little is to be gained by the use of tailored testing with conventional items for normative measurement except in the case of best and worst students.

Integrating the elements of programmed learning and sequential branched testing into a single curriculum requires a computer for electronic storage and retrieval of the material to be learned, the test items for measuring mastery, and the instructional branching strategy for both the curriculum and the tests. Suitable multi-media teaching terminals

with visual display devices, light pens, audio units, and typewriters under either student or computer control, depending upon the nature of the curriculum and purpose of the student, must be provided in large numbers at reasonable cost before computer-assisted instruction, testing, and guidance can become operational. Several major companies are now designing hardware configurations that will soon have the required functional capabilities for fully implementing computer-assisted instruction. It is now fairly certain that the cost of such a system can be sharply reduced by mass-production to the point where it is economically feasible to think of large-scale implementation (Alpert and Bitzer, 1970). Psychological laboratories for computer-assisted instruction at Stanford, Texas, Illinois, Florida State, System Development Corporation, the Mitre Corporation, and a dozen other universities and research institutes have already demonstrated the feasibility of this new technology as well as its dramatic impact upon individual learning in many areas.

Such new technologies as Project PLAN, Individually Prescribed Instruction, and computer-assisted instruction are highly promising in their eventual impact upon educational practices and the concommitant measurement of standardized mastery using criterion-referenced tests instead of normative testing for competitive selection. Successful prototypes have been developed, but these represent only a small beginning compared to what must be done in the way of research and development before individualized instruction in the true sense of the term can be properly implemented on a large scale.

National Assessment of Educational Change

Still another important departure from standardized normative measurement of individual differences in mental abilities grows out of the increased

concern for developing a national system of social indicators, measures that reflect the quality of life, the rate of educational progress, and the value of human resources for the nation as a whole as well as for different regional, ethnic, and socioeconomic groups. A recent report of the Behavioral and Social Sciences Survey Committee (1969) published by the National Academy of Sciences has recommended the establishment of a system of social indicators by the federal government which would lead to an annual social report for measuring changes in many aspects of society. A step in this direction has already been taken by the National Assessment of Educational Progress, a project of the Education Commission of the States (Womer, 1970).

Under the leadership of Ralph Tyler and support from the Carnegie Corporation, the Exploratory Committee on Assessing the Progress of Education began in 1964 to collect information about the knowledge and skills held by 9, 13, and 17-year-olds and of young adults in ten subject areas taught in schools. After five years of planning and public debate as to the merits of the project, National Assessment launched its first annual survey for all four age levels in three subject areas -- Citizenship, Science, and Writing. The national sample contained a total of approximately 100,000 persons carefully chosen on a stratified random basis involving 52 sampling units from each of four geographic regions.

The first step in preparing materials for National Assessment was to determine a list of educational objectives for each subject. Using these objectives as guides, various measurement research organizations took responsibility for preparing exercises designed to assess what young people actually know. A variety of approaches -- questionnaires, interviews, observations, and performance tasks -- were employed in addition

to traditional multiple-choice and short-answer questions similar to those used in standardized mental tests.

Four important distinctions can be made between the National Assessment exercises and multiple-choice items employed in normative tests. First, the assessment exercises are designed to discover what defined segments of the nation's population can do or what they know, rather than to distribute people normatively according to measured individual differences. For example, what percentage of the 9-year-olds in the country know that most plants get most of their water directly from the soil? Or know how to report a fire? Or report that they had ever taken part in some organized civic project to help other people? Does this percentage shift significantly across different segments of the population or from one year to the next?

Second, while items in a test are summed to give a score for each individual, exercises in National Assessment are each analyzed in their own right by pooling data across individuals. For this reason, it is particularly important that the exercises be meaningful to specialist and layman alike, that they be directly related to the stated objectives, that they have high content validity. Extensive review sessions involving a variety of judges were held for every exercise retained for National Assessment.

Third, the exercises are aimed at three levels of difficulty in order to report to the American public examples of knowledges, skills, and understandings that are common to almost all American youth of a given age, examples that are common to a typical or average American youth, and examples that are common to only the most knowledgeable youth. Ideally, one-third of the exercises should be passed by 90 per cent of the population

one-third by 50 per cent, and one-third by only 10 per cent. By contrast, item-difficulty level in the typical normative test is likely to hover near the 50 per cent level or to be evenly distributed throughout the range.

And fourth, the exercises are assembled in heterogeneous packages with different sets of exercises given to different individuals on a sampling basis. A package for 17-year-olds last year, for example, contained seven multiple-choice Science exercises, three free-response Citizenship exercises, and one essay exercise for Writing. Exercises are packaged in any convenient fashion that adds up to no more than 50 minutes of assessment time for each person. Items in a normative test, on the other hand, are assembled in relatively homogeneous scales so that they can be added together to give a reliable score.

Unlike most measurement applications in psychology and education, in National Assessment a person is never asked to record his name. Responses are clustered and analyzed by sex, age, race, region, community, and family characteristics in order to obtain census-like information about the educational progress of various segments of the population. Repeated applications in the years ahead will provide a wealth of data dealing with change over time -- data that should be useful in national planning, particularly when examined together with other social indicators.

Individuals and schools approached by National Assessment were given the option of declining to participate in order to respect their rights to privacy. Exceedingly few refused to participate under these permissive conditions, testifying to the wisdom of this policy. My own experience in soliciting the cooperation of 13,000 high school students in a probability

statewide sample (Moore and Holtzman, 1965) and in asking for the continued participation of 420 families in a longitudinal study of personality development (Holtzman, et al, 1968) has been similarly favorable. Unbiased samples can be obtained in most measurement studies without coercion of even a mild sort. National Assessment provides an exemplary model of how one should proceed in order to protect the privacy of individual participants and their freedom to decline.

Preserving the confidentiality of data is a related problem that continues to worry many thoughtful individuals. As we move into large-scale programs with extensive, centralized data banks stored in computers, the possibility of harm to an individual cannot yet be completely eliminated. The files that may do greatest damage to the individual are those which are kept secret from him but not from those who can take action affecting him. While much of the national concern expressed in recent Congressional hearings deals with personal information that psychologists are unlikely to find interesting, specific attention has been directed at potential abuses of individual privacy involving psychological test data, biographical information, and social attitudinal data typically employed in psychological research. The proper balance between protecting the individual against the misuse of information about himself and collating data to help solve major social, economic, and educational problems has not yet been achieved. On the other hand, continuation of the present highly decentralized systems will not cure present abuses of individual privacy, although it will prevent the integration of information required for future social development. As Ruggles (1969) has pointed out, the key to the problem of protecting privacy is not to depend blindly on the inefficiency which

accompanies the present situation. Properly developed centralized data banks can eventually assure greater protection for the individual while also providing essential information for basic research as well as future national planning.

One interesting solution to the problem of protecting the confidentiality of data from individual respondents is the Link system that has been devised for the national study of college student characteristics by the American Council of Education Cooperative Institutional Research Program (Astin and Boruch, 1970). Measurement data and biographical information on several hundred thousand college freshmen are collected each year as part of an ongoing educational data bank. Initially, a more or less traditional system was instituted. Two physically separate tape files were created, one containing the student's answers to research questions together with an arbitrary identification number, and a second containing only the student's name and address and the same arbitrary number. The first tape with the research data file was openly accessible for analysis. The second tape with the name and address file was locked in a vault and used only to print labels for follow-up mailings. The original questionnaires and punched cards were then destroyed.

Good as it may seem, this system still did not offer complete protection against government subpoena or unauthorized disclosure by staff members with access to both files. A third file, the Link file, was created which contained two sets of numbers, the original arbitrary identification numbers from the research data file and a completely new set of random numbers which were substituted for the original identification numbers in the second file. The final step in establishing the new

system was to deposit the new Link file at a computer facility in a foreign country with a firm agreement that the foreign facility would never release it to anyone, including the American Council on Education. Follow-up mailing tapes now have to be prepared by the foreign facility. There is no way that anyone can identify individual responses in the research file.

Such elaborate steps to guarantee the complete confidentiality of personal information in research files may seem far too expensive. Why go to this extreme when the chances are exceedingly remote that any harm could be done to an individual by using a more traditional system? The reason for foolproof data files is that the public demands it. However unlikely, there does exist the possibility of court subpoena or improper invasion of privacy when the data files and decoding files are under the control of the same organization.

Recognition of Social, Cultural, and Linguistic Variability

One of the most important changes of the past decade in the field of mental measurement as well as in society as a whole is the greatly increased respect for social, cultural, and linguistic variability among different kinds of people. Until recently, the "American way of life" was defined almost entirely by middle-class values of white, English-speaking people of largely western European origin. In general, school curricula, symbols of social status and privilege, occupations, the more highly valued life styles, and to some extent even suggested definitions of intelligence, all conformed to the dominant values of which most Americans were proud. The forgotten minorities were expected to adjust to these values if they were to enjoy the fruits of the nation. As recently as ten years ago, school principals in the Southwest often pointed

proudly to the fact that the speaking of Spanish by Mexican-American children was prohibited on their school grounds, English being the only permissible language in which to receive an education.

The emergence of Black culture, the Chicano movement, and the stirring of the American Indian as well as other forgotten groups in the wake of desegregation and civil rights legislation has forced white America to re-examine its soul. The result in the field of mental measurement has been a recognition and acceptance of cultural variability, a search for new kinds of cognitive, perceptual, and affective measures by which to gauge mental development, and a renewed determination to contribute significantly to the task of overcoming educational and intellectual deprivation.

A generation ago the typical study involving mental measurement and social variability consisted of giving tests standardized largely on middle-class whites to people of other ethnic, linguistic, and socio-economic background. Countless individual and group differences were observed and classified in a descriptive manner. Today more attention is given to devising procedures for measurement and evaluation which are indigenous to the culture under study. Illustrative of this new approach is the work of Freeberg (1970) who developed a test battery specifically tailored in content, format, and administration to disadvantaged adolescents drawn largely from the Black and Puerto Rican ghettoes of New York. The extensive six-year longitudinal study of 2000 Headstart children undertaken last year by the Educational Testing Service also contains a large variety of new measures that are specifically designed for culturally disadvantaged children (Anderson, 1969). The problem with most such tailored procedures is that they may be just as ill-suited for use with

other markedly different individuals as are tests standardized on middle-class whites when employed for assessing educationally disadvantaged children.

The most difficult methodological problems arise in cross-cultural research where two or more distinctly different cultures are compared systematically (Holtzman, 1968). The translation, calibration, and administration of psychological measures across cultures requires close and continual collaboration of specialists from each culture who have learned to trust each other fully. In a similar manner, measurement across subcultures within a given nation requires the full participation of representatives from each subculture, a condition that is met by all too few investigators thus far. In spite of such problems, studies dealing systematically with cultural, social, and linguistic variability are growing rapidly in number while also increasing greatly in the power of their research designs. Is it too much to hope that by the end of the coming decade the lingering ethnocentricism of the testing movement will disappear?

* * *

In the short span of this paper it has been possible to highlight only selected topics within the broad field of mental measurement. It should be obvious to even the casual observer of trends in the field that other areas also deserve attention. It is worth noting that every one of the new advances reviewed is heavily dependent upon the modern electronic computer for its implementation. Fundamental to the changing world of mental measurement is the rapid growth in power, versatility, and accessibility of high speed computers. Large-scale testing; new educational

technology such as individually prescribed instruction, sequential branched testing within the curriculum, Project PLAN, and computer-assisted instruction; national assessment of educational change and the development of a system of social indicators; new techniques for preserving the confidentiality of personal data; and even new programs for assessing the mental development of culturally different people -- all require a computer for implementation.

In focussing primarily upon the social implications of new advances, it is easy to overlook the numerous theoretical and methodological contributions to the field of measurement and evaluation that have been made in the past few years. New techniques of scaling, test theory, factor analysis, and multivariate experimental designs are being produced and extended in a lively manner. The immediate social significance of these developments may not be readily apparent because of their indirect, long-range nature as basic research contributions. And yet, without the continued, vigorous support of such theoretical and methodological advances, the truly great potentiality of the changing world of measurement would fail to materialize. Each of the promising new developments surveyed above is heavily dependent upon the solution of difficult basic research problems before it can be fully realized to the benefit of society. There is every reason to be optimistic about the next ten years in the field of mental measurement, given the recognized social significance of new developments and the rapid rate at which basic work is advancing.

REFERENCES

Alpert, D. & Bitzer, D. L. Advances in computer-based education. Science, 1970, 167, 1582-1590

Anastasi, A. Psychology, psychologists, and psychological testing. American Psychologist, 1967, 22, 297-306.

Anderson, S. B. The ETS-OEO longitudinal study of disadvantaged children. In Untangling the Tangled Web of Education. Princeton, N. J.: Educational Testing Service, 1969.

Astin, A. W. & Boruch, R. F. A "Link" system for assuring confidentiality of research data in longitudinal studies. ACE Research Reports, Vol. 5, No. 3. Washington, D. C.: American Council on Education, 1970, 18 pp.

Behavioral and Social Sciences Survey Committee. The behavioral and social sciences: Outlook and needs. Washington, D. C.: National Academy of Sciences, 1969.

Brim, O. G., Jr., Glass, D. C., Neulinger, J., Firestone, I. J. & Lerner, S. C. American beliefs and attitudes about intelligence. New York: Russell Sage Foundation, 1969.

Commission on Tests. ~~Final~~ Report of Commission on Tests: I Righting the Balance, New York: College Entrance Examination Board, 1970.

Cooley, W. W. & Glaser, R. The computer and individualized instruction. Science, 1969, 166, 574-582.

Dunn, J. A. The accommodation of individual differences in the development of personal programs of study. Paper read at symposium, "Project PLAN: a computer-supported individualized education program." Annual meeting of the American Psychological Association, Washington, D. C., 1969.

Ferguson, R. L. Computer-assisted criterion-referenced testing. Working Paper No. 49, Learning Research and Development Center, University of Pittsburgh, 1969.

Freeberg, N. E. Assessment of disadvantaged adolescents: A different approach to research and evaluation measures. Journal of Educational Psychology, 1970, 61, 229-240.

Holtzman, W. H. Cross-cultural studies in psychology. International Journal of Psychology, 1968, 3, 83-91

Holtzman, W. H., Diaz-Guerrero, R., Swartz, J. D. & Lara Tapia, L. Cross-cultural longitudinal research on child development: Studies of American and Mexican school children. In J. P. Hill (Ed.) Minnesota Symposia on Child Psychology, Vol. II. Minneapolis: University of Minnesota Press, 1968.

Holtzman, W. H. Computers in education. In Holtzman, W. H. (Ed.) Computer-Assisted Instruction, Testing, and Guidance. New York: Harper & Row, 1970.

Lord, F. Some test theory for tailored testing. In Holtzman, W. H. (Ed.) Computer-Assisted Instruction, Testing, and Guidance. New York: Harper & Row, 1970.

Ruggles, R. How a data bank might operate. Think, 1969, 35, No. 3, 22-23.

Womer, F. G. What is National Assessment? Denver, Colorado: Education Commission of the States, 1970.